Situated Conversation with a Communicative Interface Robot

Katashi Nagao
Dept. of Information Engineering
Nagoya University
and CREST, JST
Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan
nagao@nuie.nagoya-u.ac.jp

Abstract

This paper will present a new approach to natural language communication with an interface robot that is aware of real world situations such as location and time. The robot also memorizes individual users and personalizes conversation with each user using experiences of past conversation with the user and his/her information-seeking behavior.

Introduction

It is expected that in this century, the style of interaction between human and machine would change dramatically because of the remarkable progress of the robotic technology and other emerging technologies. This project will apply the new, integrated and general protocol and platform for the new interaction style between human and machine and we are aiming at achieving the new considerable market position through them. Considering the new style of the relationship between human and machine, we will employ the concept of spoken conversation as a core interaction method, instead of the current one-way operation. In order to produce the conversation between man and machine, we need to construct a new, universal protocol and platform for natural language processing. This project also provides the emotion unit. It will be a core program of conversation system. This resulting system is expected to be the universal tool for various existing information terminals. Furthermore the easier interaction will obtain a large amount of potential customers.

Existing conversational systems such as natural language interfaces are constrained in detail and the conversation is restricted and not natural. In these systems, the user has to say the command with absolute precision and the machine cannot answer to the unexpected sentence. Moreover the context is tightly bounded. Because of these problems, conversations in the systems converges toward one purpose. This is the very limited aspect of daily conversations of human beings, therefore it sounds unnatural for users. That is the one of the reasons why voice conversation systems still remain in uncommon. The systems are used mainly for the service like telephone inquiry operator.

If you observe daily conversation, you will find that the various contents develop in the various ways, and that the multi-goal oriented speech acts appear simultaneously. It is obvious the coherence between one utterance and another is so weak in the conversation. Computers must be adjusted to such characteristics on the conversation to make the conversation system fit for practical use. Daily conversations always develop unexpected way by nature. The proposed conversation system is expected to deal with such interactions and make the atmosphere of natural conversation.

The proposed system consists of Utterance Reactor, Plan-based controller and Emotion Unit. Utterance Reactor outputs information for inputs from local points of view. On the other side, Plan-based controller controls conversations in terms of global phase of conversations (Carberry 1990). And the design of Emotional Unit is that the former controls the latter and the latter activates the former. Emotion Units get information from both and generates autonomous emotional states. And the generated emotional data will be provided to the each conversation systems. By introducing this flexible discourse navigation unit, interactions more closer to conversations of human-beings must be available.

We emphasize the spoken language technology instead of the present written language based technology. The spoken language technology is based on the discourse structure marker by phrases or words whereas the conventional natural language processing based on sentence. This technology makes it easier to deal with the characteristics of spontaneous speech such as ungrammatical sentence, slip of the tongue and restating, disfluency and filler, and the problem of voice recognition.

Communicative Interface Robots

Based on the architecture, as a new human interface, we have been promoting a research on an interactive robot that is able to have a conversation with human beings. The robot that we are working on is not a robot whose appearance and function like the movements of arms and legs, are being pursed. We have focused on the mechanism that the robot should be helpful and useful for human beings, as a sort of assistant. In short, we are aiming at the intellectual part of the robot. The means of communicating with a human being is conversation, which is very familiar to us. However, unless we overcome the issue that there is a resistance to the action of human's talking to a machine, it won't be successful. It seems that once a human is unintentionally sensing the existence of an object that is giving notice to him, he

is likely to intuitively aware that, “This is the object with which I communicate.” In other words, with an intuitive and realistic conversation, in order to enable a smooth communication between a human being and a machine, we utilize robots as such containers. One of them is a robot whose name is Pong, being introduced as follows.

Pong

When Pong is spoken to, it understands user’s intention, collects the necessary information, processes it, and tries to create an adequate response. Additionally, Pong will voluntarily perform an action preferable to the user, by being equipped with an ability that memorizes the history of conversations. Also this robot will express happiness in response to the user’s happiness and will be delighted to know the user’s joy, too, if a system with an ability of comprehending human’s emotion, like specializing and attuning the user, is pre-built in. When this robot comes to possess both intellect and emotion, it will play the roll of a healer as well as convenience for humans.

Figure 1 shows a snapshot of dialogue between Pong and the author.

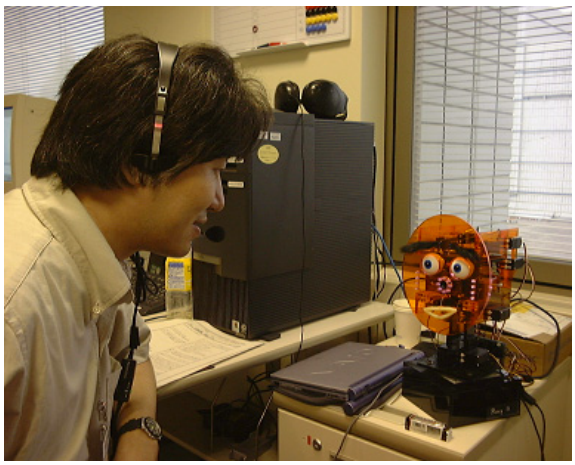

Figure 1: Dialogue Snapshot with Pong

Lately, as a pet type robot appeared, concern over the robot that manages a certain type of conversation with humans has increased. The pursuit of entertainment alone will result in a human’s getting bored of a robot someday, however, with intelligence of processing dialogue with humans, and knowledge, it can become a truly reliable partner to them. The following is an example dialogue between Pong and a human (actually the conversation is done in Japanese).

Pong:	Hello. I am Pong, born at IBM Almaden Research Center in the States. I understand a little about human languages. Ask me something.
Human:	What can you do?
Pong:	Various things, like calculation.
Human:	Well, here is a question.
Pong:	All right.
Human:	Multiply 100 by 200?
Pong:	The answer is 20,000.
Human:	How about dividing 100 by 7?
Pong:	It isn’t divisible. 14.28571428 ...
Human:	Thank you!
Pong:	Any time.

Named Pong after Ping Pong, this robot was originally developed at IBM Almaden Research Center. It is a product of BlueEyes Project (IBM Almaden Research Center 2002), which has been aiming to develop a computer that comprehends human’s emotion.

Equipped with infrared LED in its cheeks, the robot inputs human’s image by using the camera installed in its nose. It can show reflective responses to human’s behavior such as smiling toward a human being and expressing its sadness when losing sight of him. What is important here is that the facial expression activates a communication with a human being.

The same is equally true of computer animation. Added the character look and given expressions or gestures what are called non-verbal modality, several studies were made. Nowadays, on hearing the name of interface agent, quite a few might well imagine personified characters on the computer screen. But what is absolutely deficient in its character on the screen should be some kind of presence or awareness, which describes “He/she exists there now and is concerned about me.” Nothing is more appealing to human’s intuition than physical presence. The aim of virtual reality is exactly the same as it, that is, to make that part controllable by software.

It can be possible to eradicate an image that computer including its circumstances is hard to approach and use, when interface comes to possess physical presence, to have dialogue with human, and to work on him, strengthening friendship with human.

Such a concept enabled to produce a robot like Pong. We can get news and information of that day not by watching TV and reading newspaper but by talking to the robot. An example is as follows:

Human:	Pong, do you know Internet?
Pong:	Of course, I know it thoroughly. We can get various kinds of information with Internet.
Human:	Let's see ... I want to see the page of Nikkei.
Pong:	A homepage of Nihon Keizai Shinbun, right? Here comes in the screen.
Human:	Can you read it for me?
Pong:	On the Tokyo Stock Exchange ...
Human:	I'd like to see homepages in English language. Please show me a Web page of BlueEyes.
Pong:	Here comes the BlueEyes' Web page.
Human:	Do you understand English?
Pong:	Sure thing, although not very well yet.
Human:	Can you explain it?
Pong:	I am explaining what the homepage says. BlueEyes is ...

It is expected that the time of information appliance will come in the future, every home electric appliance being computerized and networked. When home electric alliances are networked, a new problem will occur such as how to use them. Currently, we need simple operation, for instance, when we want to watch TV, we just need to switch it. In brief, there is only a simple relation between doing and being done now. However, if various machines have dependency relations each other, we will be confused at which operation effects what.

Since it is troublesome for a human being to be familiarized with these dependency relations, we expect the robot as interface to undertake a roll of human. Inputted all the functions and presence relations, when the word of human's demand is conveyed to it, the robot will recognize both the intention and situation of a human being, operates home electronic alliances, and reports the circumstances. Thus only after this happens, we believe a robot will become a dispensable asset for human beings.

In addition to speech recognition, voice synthesis and dialogue processing technology, Pong is integrated with the technology of digital content what we call *semantic transcoding* (Nagao, Shirai, & Squire 2001). This way a robot for dialogue can be a test-bed integrated with various technologies.

Finally I am going to describe the future direction for interface agents and robots. The weak point of robot being physical asset is that it cannot move in the information world as what it is. As the memory and knowledge of robot is significant for humans, still there will be a number of occasions that people want to use it even when they are out of home. It is imaginable that a human being is rambling with it. However, it will be more practical to make the interaction possible while moving by means of residing software, that is interface agent, in the portable system, which carrying its memory and knowledge.

In this case, we won't feel its physical presence. None the less with its previous remembrance of interaction it will enable us to sense that the relation of the robot and agent is consistent. Then we will be able to send a request, by a conversation with the agent as if we talk to a faraway person.

In this manner, it seems that an interface agent and a conversational robot for dialogue will share as much as theirs, will be consistent for their memory and will change their style of interaction depending on conditions.

Also the robot and the agent will become the medium of community and will function as a system of activating communication between humans. For instance, a robot for dialogue makes it possible for people far apart to experience a face-to-face meeting when communicating asynchronously. Continuously, a robot or an agent will come to act as a system of possessing externalized memories or personality of an individual, throughout its long experience of gaining a personal information. This implies that a robot or an agent can become an agent for a human being literally. Suppose such an issue as credibility and responsibility is settled technically and socially, a human being will be able to make his own copy or his doppelganger and will make it possible to do multiple tasks at a time.

QB

QB is a mobile robot for a dialogue with an expressive appearance and will become a successor machine. This robot can respond to various questions listening to human languages as well as Pong. It can look up necessary information from Web content and convey it to a human being with voice like the weather of next day, today's news and the result of a sporting event. Additionally QB has a human-like face and its expression provides a conversation with reality and affinity. Moreover, when we touch its head or hand, sometimes it enjoys it and sometimes not. This happens because it is trying to pay the user's attention emotionally.

Figure 2 presents a scene of conversation between QB and the author. Different from Pong, as QB is able to move autonomously, a human being uses PDA (Personal Digital Assistant) which possesses wireless function to talk to QB. And he can send the user's required ID to be confirmed and his present location.

Figure 2: Dialogue Snapshot with QB

Also QB tries to change its location at its discretion and to have a conversation with a person in the distance. For

example, when it is called with voice from far away, QB approaches nearby and starts to talk. Thus QB has an ability of recognizing the location of both QB and the user and of moving autonomously. Needless to say, when it almost bangs into something, QB can avert it and make its way around.

QB has its own house where it can recharge its battery. Accordingly, if its energy is getting low, QB comes back to its house. The house is equipped with function of processing information to match an individual demand and we can exchange information with QB by radio transmission.

Furthermore, listening to a number of voices, QB makes an adequate reply depending on the individual. Like a walking bulletin board, it can play the role of a messenger who collects various opinions and distributes them on to others. In the future, we can expect a robot like QB will become popular among a lot of people and will activate communication. Eventually, it will be an intelligent partner for a human being, answering a human's question casually.

The main objectives for QB are the following:

- As the next generation of user's interface, we propose an interactive robot with both functions of voice-conversation and communication. Also we manufacture and experiment it.
- The robot stimulates a user's interest, by its physical presence and diversely personified expressions.
- With its response to adapted to various contexts, it accelerates the user's continuous usage.
- The robot interacts with the server and offers a user various information service through the Internet, taking care the complicated procedures for the user.
- The robot remembers a user, changes the response depending on his traits, and personalizes the content.

Situated Conversation

QB can move from place to place autonomously and change situation of conversation. When a new situation is introduced, it automatically restricts a context of dialogue according to the situation.

Spoken conversation is usually based, not only on linguistic contexts, but also on non-linguistic contexts relating to a real world situation. This is called *situated conversation* (Nagao 1998). In situated conversation, the topic and focus of speech depend on the situation and are easily recognizable when the participants are aware of their surroundings. Our robots can handle situated conversation by virtue of the basic properties of communicative interface robots.

Knowing the user's intention is necessary for natural human-robot interaction. Although a real world situation would provide just a clue about the intention, being able to integrate the non-linguistic context introduced with that situation, with the linguistic context constructed by dialogue processing, is an important step forward.

For instance, we have developed some methods for situation awareness such as location detection techniques using GPS and ultrasound sensors and object detection techniques using object IDs.

Situation Awareness of the Real World

Situation detection is classified into ID-based and location-based methods. ID-based methods (called ID-awareness) mark real world objects with machine-recognizable IDs (e.g., barcodes, infrared rays, radio waves). Recognition of objects can be extended to the recognition of situations. Suppose that there is an ID on every door in a building. When the user stands in front of a door, the mobile system detects the location by scanning the ID on the door and by processing the information related to this position may derive some understanding of what the user intends to do. Location-based methods (called location-awareness) include GPS, three-dimensional electromagnetic sensors, gyroscopic sensors, and so forth. Spatial information is also a useful input for attaining situation awareness. In contrast to ID-awareness, location-awareness is more scalable, because it doesn't require that objects be tagged. However, when the location of physical objects changes, the system has no way of recognizing this movement and so will fail to identify and call them to the user's attention properly. Therefore, it would be better to apply a hybrid approach that uses both ID-awareness and location-awareness to complement each other.

Based on ubiquitous computing (Weiser 1993), recognizing the human environment will become easier, because it proposes that very small computational devices (i.e., ubiquitous computers) be embedded and integrated into the physical environment in such a way that they operate smoothly and almost transparently. These devices are aware of their physical surroundings and when a human uses a physical device that contains ubiquitous computers or enters some area where physically-embedded computers are invoked to work, these computers are aware of the human's activities. From the viewpoint of reliability and cost-performance, ubiquitous computing does not compare well with mobile computing, since ubiquitous computers require very long battery lives and are significantly difficult to maintain. In addition, when ubiquitous computers are personalized to users, as in an active badge system (Want *et al.* 1992), for example, all user personal data is processed in the shared environment, while in mobile computing, each user's personal information can be encapsulated within their own machine. Thus, ubiquitous computing also experiences privacy problems.

Situated Conversation with an Interface Robot

We have implemented some situation-awareness methods in QB and the real world environment. When a user in a library, QB interacts with the user and guides information about books in the library. The following is an example dialogue.

QB: Here is the library. May I help you.
Human: I am looking for books on computer science.
QB: (Move to the computer science bookself) Here is the computer science bookself. There is a new book on digital content. It will be helpful for your research.
Human: Thanks. But today I want to study programming.
QB: Which programming language are you using?
Human: Recently I started programming in Java.
QB: On the third shelf from top there are some books on Java for beginners.

QB detects the physical situations from RF (radio frequency) ID markers on objects such bookshelves, books, and user devices.

QB's conversational plan inference is initially triggered by introducing a new non-linguistic context, since the motivation of our situated conversation is closely related to the physical actions that accompany a new situation. For example, in a situation where a person is standing in front of a bookshelf, the situation will motivate the person to search for a book, pick it up, and read it. When the system detects the current situation from the bookshelf ID, a situation-dependent plan library is introduced and used for inference.

Concluding Remarks

Through the interface robots Pong and QB, it seems that the concept of a robot for dialogue has become more accurate. Also, from the experience of developing and experimenting these prototype systems, I have realized a strong possibility that a robot for dialogue as a new user's interface, will replace personal computers and will come into wide use in general families.

From now on, as I am going both to enhance the function of robot and to propose several applications for emphasizing the difference between interface robots and user interfaces on PCs, I plan to discuss its effectiveness.

Acknowledgments

I greatly appreciate the assistance provided by people contributed to this research. Student researchers, Shingo Hosoya, Yoshinari Shirai, Shigeki Ohira, Ryuichiro Higashinaka, Mitsuhiro Yoneoka, Daisuke Ito, Kevin Squire, Takeshi Umezawa, Toshiki Fukuoka, Yukiko Katagiri, Miki Saito, and Takashi Oguma, whom I have worked with during my time at IBM supported me to develop prototype systems.

References

Carberry, S. 1990. *Plan Recognition in Natural Language Dialogue*. The MIT Press.

IBM Almaden Research Center. 2002. BlueEyes Project Web Page. http://www.almaden.ibm.com/cs/blueeyes/.

Nagao, K.; Shirai, Y.; and Squire, K. 2001. Semantic annotation and transcoding: Making Web content more accessible. *IEEE MultiMedia Special Issue on Web Engineering* 8(2):69–81.

Nagao, K. 1998. Agent augmented reality: Agents integrate the real world with cyberspace. In Ishida, T., ed., *Community Computing: Collaboration over Global Information Networks*. John Wiley & Sons.

Want, R.; Hopper, A.; Falcao, V.; and Gibbons, J. 1992. The active badge location system. *ACM Transactions on Information Systems* 10(1):91–102.

Weiser, M. 1993. Some computer science issues in ubiquitous computing. *Communications of the ACM* 36(7):74–85.